report.md

Task1

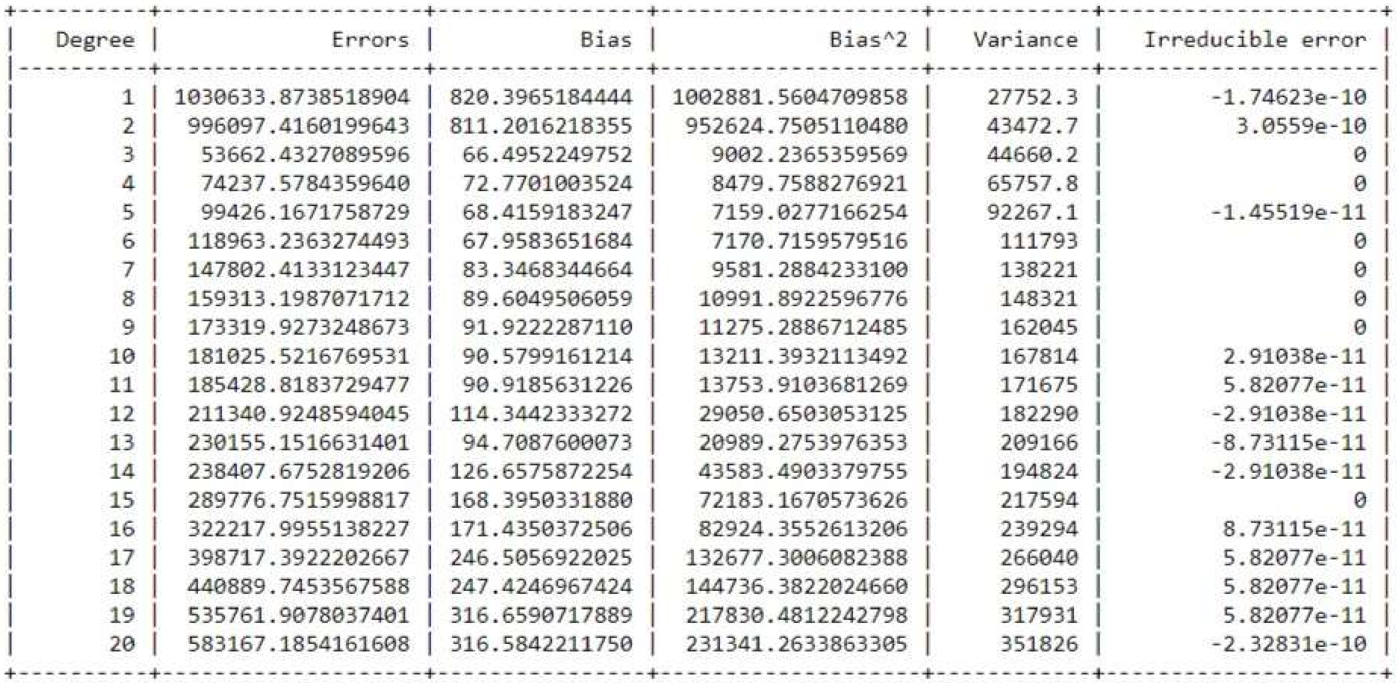
Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. This function takes 2 variables let say first variable X and second variable Y where Y is column matrix(vector) of order n and X is a (m+1)*n matrix which gives m cofficents(a_1 to a_m) and 1 intercept (a_0) in a hypothesis function

$$h = a_0 + a_1*b_1 + a_2*b_2+ ... + a_{m-1}*b_{m-1}+ a_m*b_m$$

where $b_i = x^i$ converted by PolynomialFeatures().fit_transform() function.
It fits multiple lines on data points and returns the best-fit regression line. To achieve this line, the model derived using LinearRegression().fit() aims to minimize the Cost Function(J) [The Root Mean Squared Error(RMSE) between predicted Y value (y_pred) and true Y value (y)]. And we can get the predicted values for the test set using LinearRegression.predict() which evaluates y_pridicted for x in the test set by using derived hypothesis function.

**X,Y are matrix and x_i ,y are elements of matrix*

Table



```
+-----------+--------------------+----------------+--------------------+----------+-------------------+
|    Degree |             Errors |           Bias |             Bias^2 | Variance | Irreducible error |
|-----------+--------------------+----------------+--------------------+----------+-------------------|
|         1 | 1030633.8738518904 | 820.3965184444 | 1002881.5604709858 |  27752.3 |      -1.74623e-10 |
|         2 |  996097.4160199643 | 811.2016218355 |  952624.7505110480 |  43472.7 |        3.0559e-10 |
|         3 |   53662.4327089596 |  66.4952249752 |    9002.2365359569 |  44660.2 |                 0 |
|         4 |   74237.5784359640 |  72.7701003524 |    8479.7588276921 |  65757.8 |                 0 |
|         5 |   99426.1671758729 |  68.4159183247 |    7159.0277166254 |  92267.1 |      -1.45519e-11 |
|         6 |  118963.2363274493 |  67.9583651684 |    7170.7159579516 |   111793 |                 0 |
|         7 |  147802.4133123447 |  83.3468344664 |    9581.2884233100 |   138221 |                 0 |
|         8 |  159313.1987071712 |  89.6049506059 |   10991.8922596776 |   148321 |                 0 |
|         9 |  173319.9273248673 |  91.9222287110 |   11275.2886712485 |   162045 |                 0 |
|        10 |  181025.5216769531 |  90.5799161214 |   13211.3932113492 |   167814 |       2.91038e-11 |
|        11 |  185428.8183729477 |  90.9185631226 |   13753.9103681269 |   171675 |       5.82077e-11 |
|        12 |  211340.9248594045 | 114.3442333272 |   29050.6503053125 |   182290 |      -2.91038e-11 |
|        13 |  230155.1516631401 |  94.7087600073 |   20989.2753976353 |   209166 |      -8.73115e-11 |
|        14 |  238407.6752819206 | 126.6575872254 |   43583.4903379755 |   194824 |      -2.91038e-11 |
|        15 |  289776.7515998817 | 168.3950331880 |   72183.1670573626 |   217594 |                 0 |
|        16 |  322217.9955138227 | 171.4350372506 |   82924.3552613206 |   239294 |       8.73115e-11 |
|        17 |  398717.3922202667 | 246.5056922025 |  132677.3006082388 |   266040 |       5.82077e-11 |
|        18 |  440889.7453567588 | 247.4246967424 |  144736.3822024660 |   296153 |       5.82077e-11 |
|        19 |  535761.9078037401 | 316.6590717889 |  217830.4812242798 |   317931 |       5.82077e-11 |
|        20 |  583167.1854161608 | 316.5842211750 |  231341.2633863305 |   351826 |      -2.32831e-10 |
+-----------+--------------------+----------------+--------------------+----------+-------------------+
```

Task2

Bias : Bias is the difference between the average prediction and the true value.Initially when the degree of polynomial is less, function is too simple and such a simple model can not fit our training model, hence a high bias. As the degree of hypothesis polynomial increases, the function becomes more flexible, allowing it to better mould itself to fit the training dataset. Hence, the error on the test dataset also decreases and both test and training accuracy increases therefore bias decreases. But eventually error starts increasing after polynomial of degree 3 which hints at the fact that the data is best modelled by a degree 3 hypothesis, and a further increase in degree of the hypothesis would result in overfitted model with training accuracy too high but test accuracy is too bad resulting in increasing the bias again.

Variance : Variance measures the spread of the prediction, which is the variability of the prediction. The variance of the models shows a general increase with an increase in the degree of the hypothesis. This is because as the degree of the polynomial increases and it becomes more flexible, it also becomes more susceptible to minor variations in the training dataset. Hence, each time the model is trained, the increased flexibility of the higher degree polynomials causes the coefficients to turn out significantly different due to differences in the training set. Hence, the high variance on the test dataset.

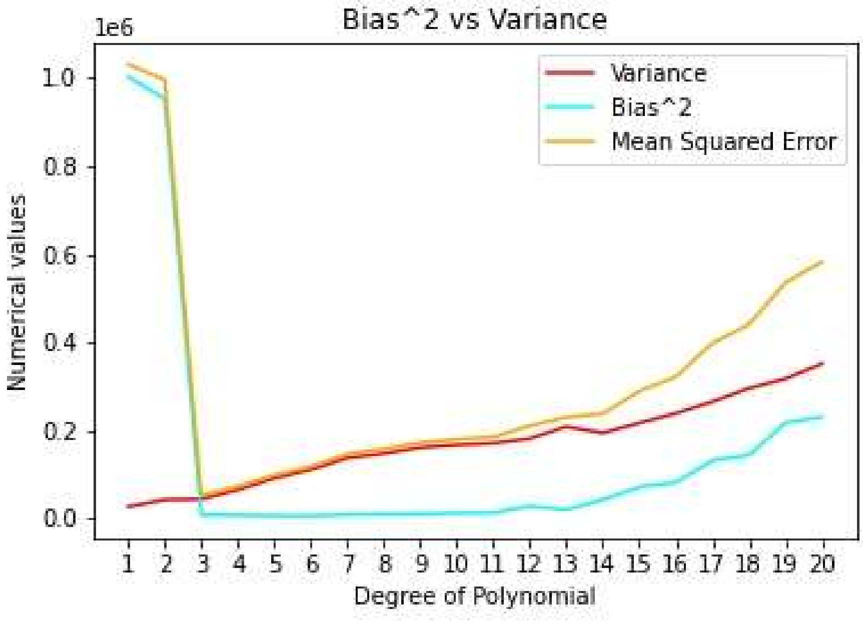
Task3

Irreducible error does not follow any general trend on increasing the degree of polynomial, it is quite random as it is continously fluctuating in magnitude as well as sign but always confined in order of 10^{-10} or even less than this and sometimes ends up to `0` . There can be multiple reasons for so

- Due to less noises in given dataset.
- Due to randomness in given dataset due to random spliting of training dataset.
- Since irreducible error (Var(noice)) can not be a negative but negative values in our output can be due to the insistence on unbiased estimators.
- Due to floating point precision error while taking mean multiple times.

Irreducible error depends more on the dataset which decides its value than the degree of hypothesis polynomial(complexity of model).

Task4



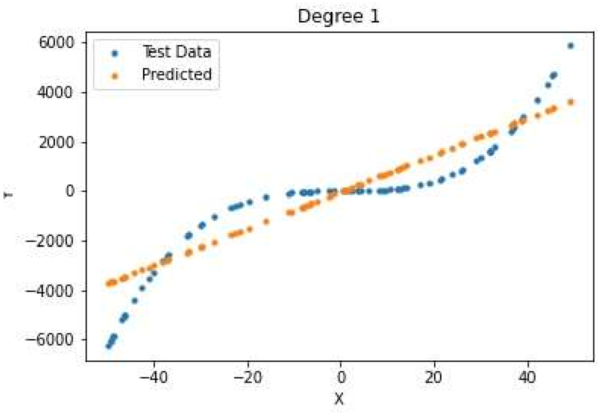
From the above graphs and tabulated values, we observe that with an increase in the complexity of the model the bias first decreases and then increases whereas variance continously increases.

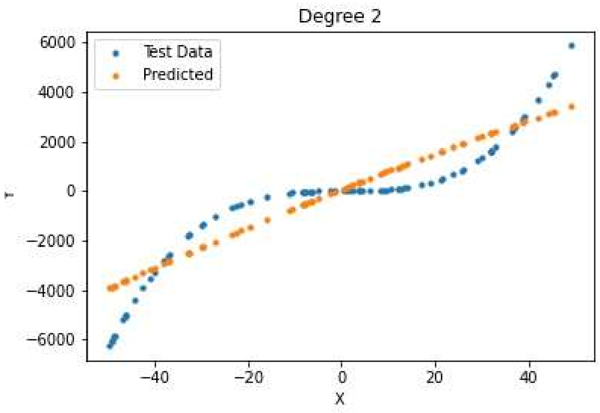
Overfitting : The phenomenon of memorization can cause overfitting. That is with increase in the number of features, or complexity/flexibility of the model (essentially increasing the degree of the model to best fit the training data) the model extracts more information from the training sets and works well with them. However at the same time it will not help us generalize data and derive patterns from them. Thus the model may perform poorly on test data sets. This is reflected in the increase in variance with increase in complexity. Thus the model is said to be overfitting. Higher degree polynomials can usually be overfitting.

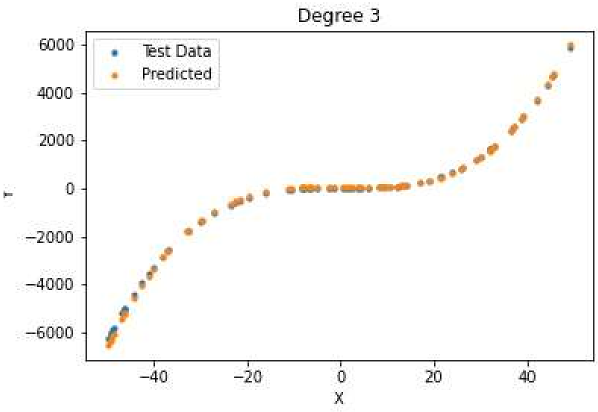
Underfitting: When a model is underfit, it does not perform well on the training sets and does not perform well on the test sets either. That is it fails to capture the underlying trend of the date. Furthermore for lower degree polynomial models, we observe a high bias but low variance. The reason being, the model may not be able to perform well even on existing training data since the lower degree polynomials are unable to capture all features of the training data. Yet the variance is high, since the model is consistently performing poorly. Therefore lower degree polynomials can usually be underfitting.

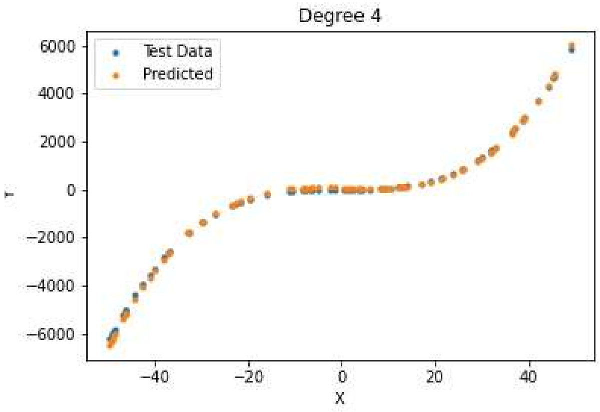
Initially when degree of polynomial is small (1 and 2) ,the model is underfitted and hence high bias and low variance and then suddenly bias and error decreases when polynomial degree becomes 3 indicating that this is the best fit case . after this , model started becoming more and more complex resulting in overfitting (which results in high variance as compared to that of smaller degrees).Over-fiiting and Under-fitting can also be observed by analyzing the error plot in the graph . Firstly , error is high when the model is under-fit and it decreases suddenly when it becomes best-fit and then again it starts increasing when it becomes over-fit.

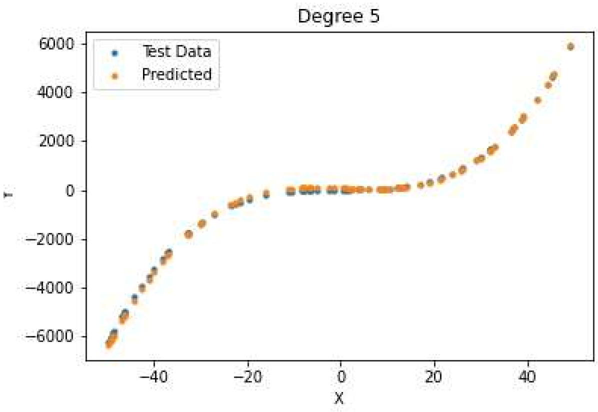
Graphs

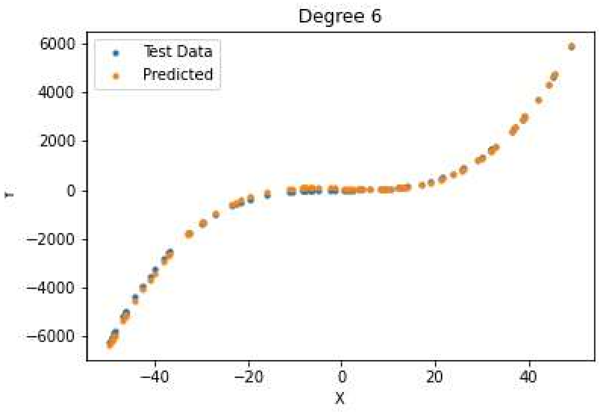

Degree 1

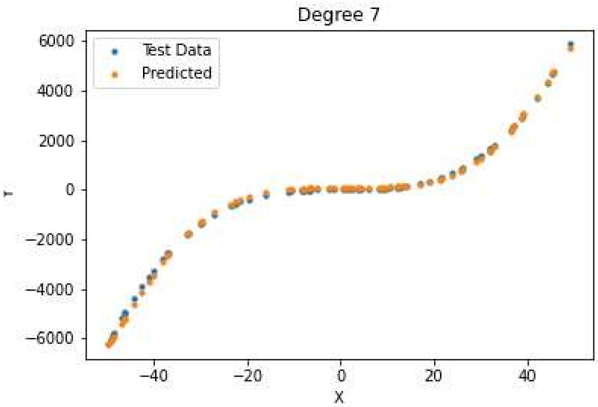

Degree 2

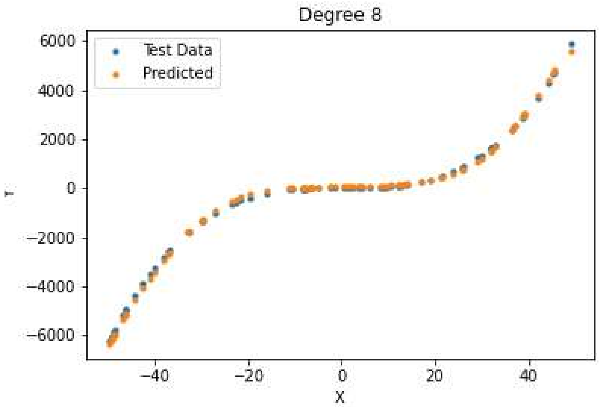

Degree 3

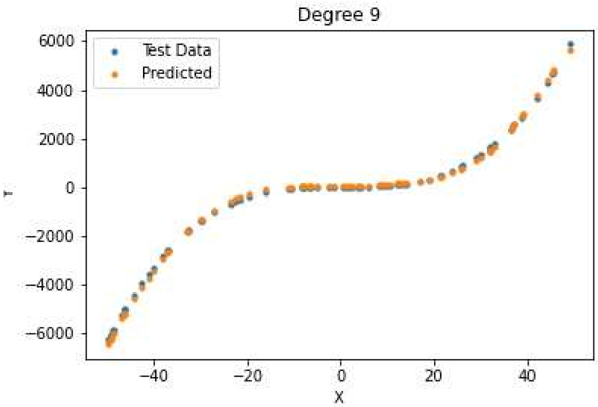

Degree 4

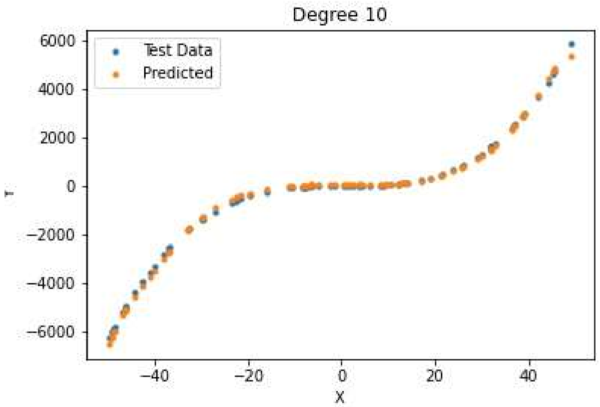

Degree 5

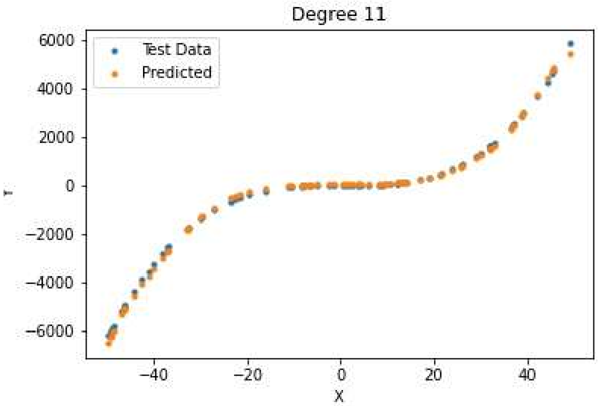

Degree 6

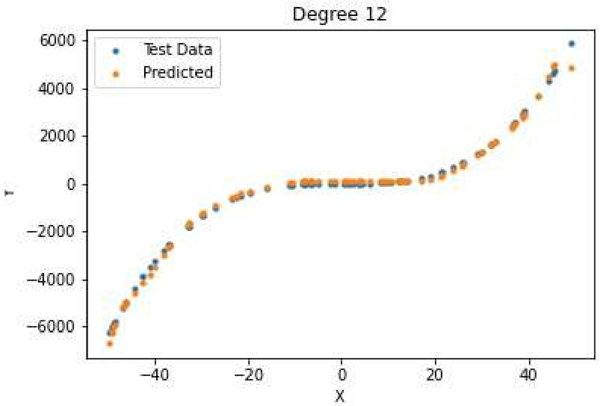

Degree 7
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

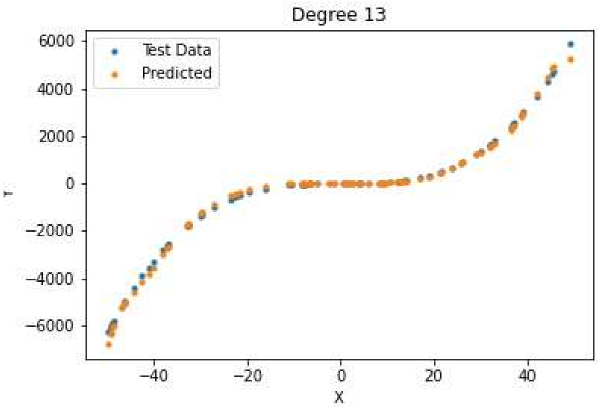

Degree 8
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

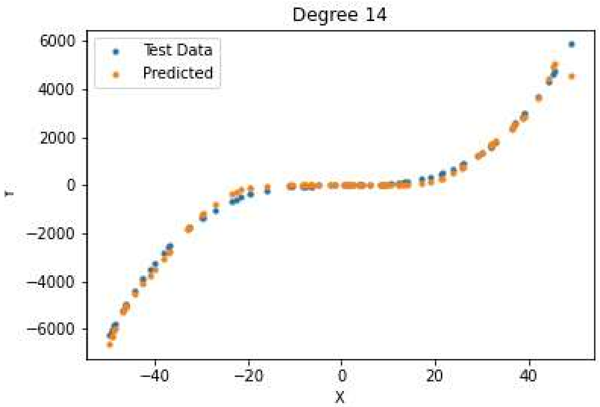

Degree 9
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

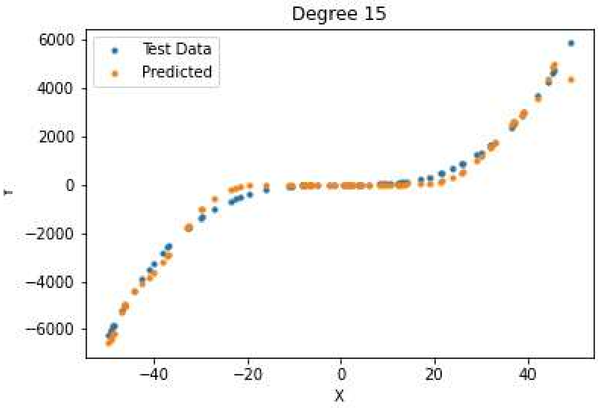

Degree 10
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

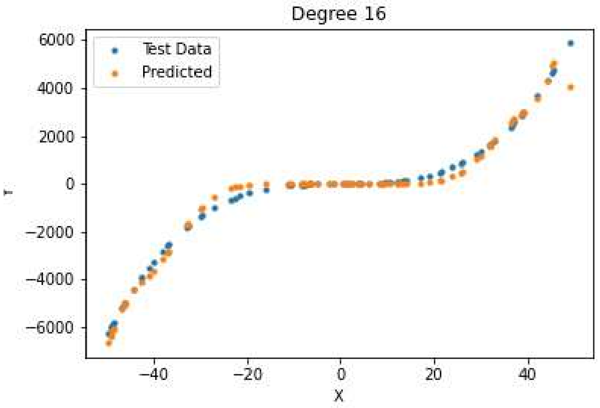

Degree 11
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

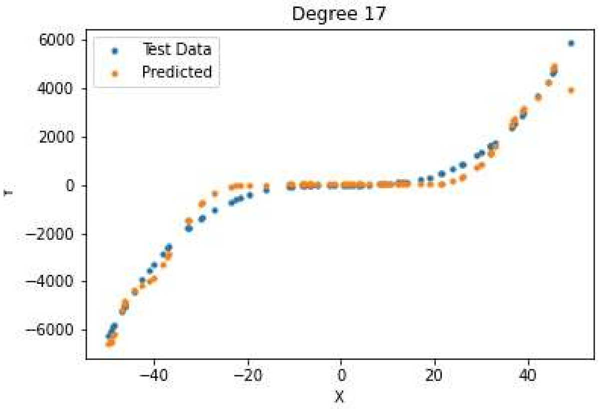

Degree 12
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

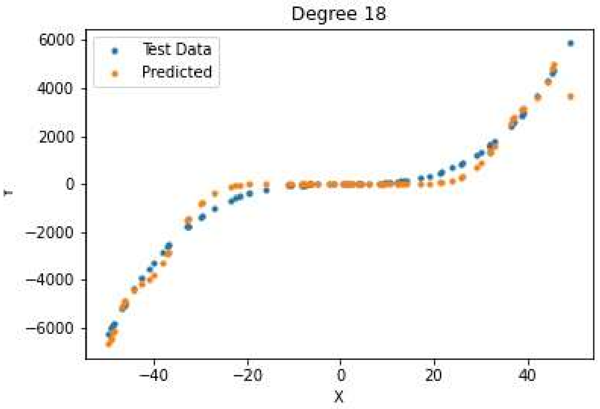

Degree 13
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

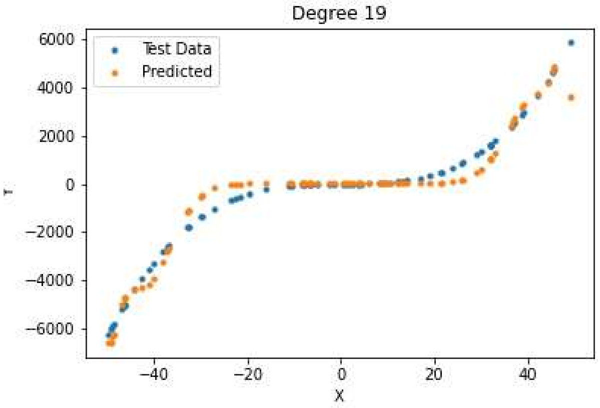

Degree 14
Test Data
Predicted
6000
4000
2000
0
−2000
−4000
−6000
Y
−40
−20
0
20
40
X

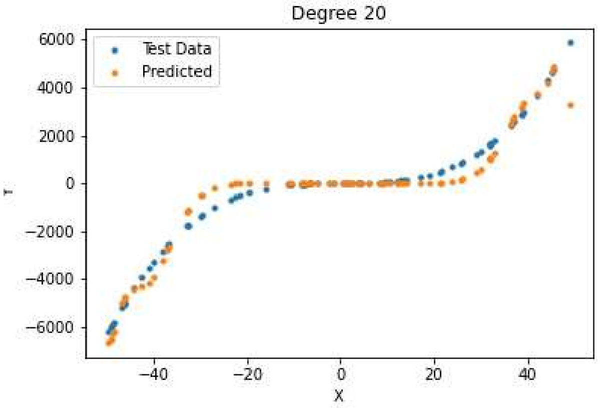












These plots show the performance of each of the 20 polynomials on the test dataset. Because we have 10 randomly distributed model of trainig set, we have used the average values of all 10 models for a data point in test set. These plots clearly show the trends that have been described above. As the degree of the hypothesis polynomial increases, the output more closely follows the actual values indicating a lower error, or bias on the test as well as training dataset. However, the error starts increasing once the degree crosses 3, which indicates that the optimal degree of the hypothesis would be around 3.The squared bias decreases drastically in transition from degree 2 to degree 3 polynomial. This is because of the nature of the test data, It looks very similar to a cubic equation. Hence , we can say that the data set resembles a polynomial of degree 3 as at this point , model is neither over-fitting nor under-fitting.One can also notice how minor variations in the dataset affect the higher degree polynomials. Although the plots do not explicitly show it, this can easily be extrapolated to indicate a high variance in those models.